

# Three Big Trends

**Explosion of Data** 

- > 90% Unstructured
- > Video & Image Content
- Needs Higher Throughput & Real-Time Computing

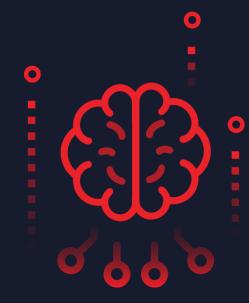




# Three Big Trends

Dawn of Al

- > Adoption Across All Industries
- > Injecting New Intelligence into Apps
- > From Endpoints to Edge to Cloud





# Three Big Trends

Computing After Moore's Law

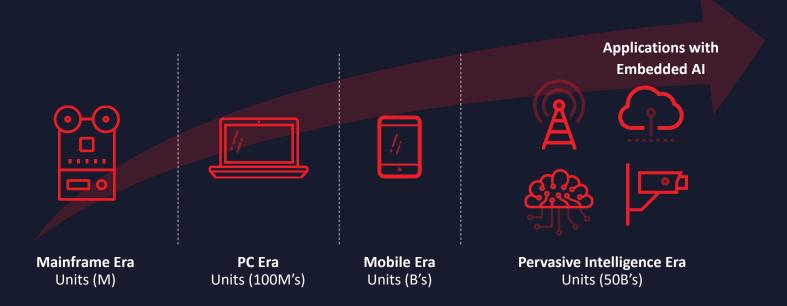
- Heterogeneous Computing with Accelerators
- > Breadth of Apps and AI Require Different Architectures
- Speed of Innovation Outpacing Silicon Design Cycles





### Era of Pervasive Connected Intelligence

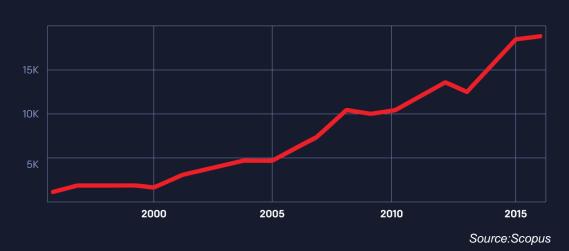
Adaptable Platforms Enable Rapid Deployment of Domain Specific Architectures (DSA's) for Performance and Energy Efficiency





# The Imperative to Accelerate Innovation Deployment to Market

### **Al Papers Published**



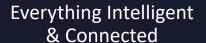
### **Over-the-Air Updates**





# The Era of Pervasive Intelligence Needs Adaptable Computing and DSA's







Deployed at Massive Scale

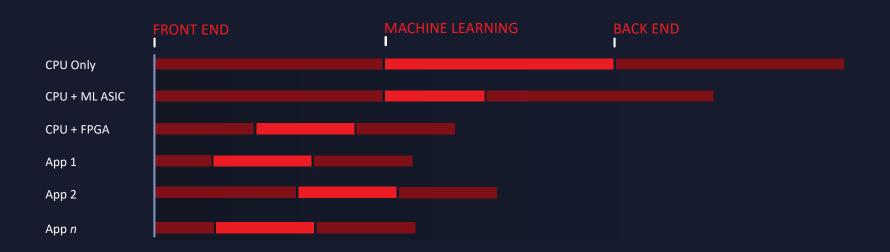


Dynamic Markets Need Rapid Innovation & Deployment



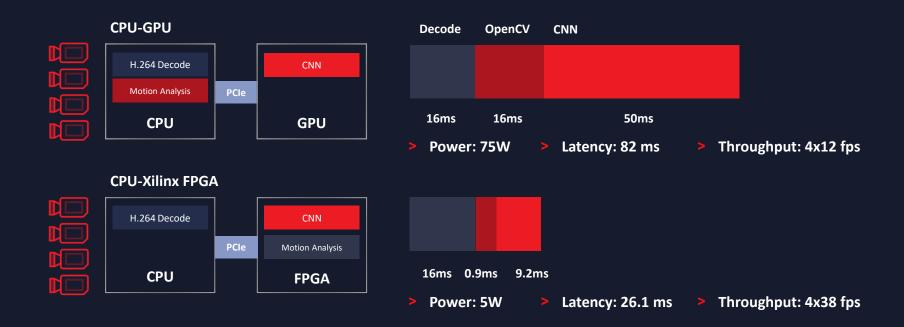
# Whole Application Acceleration on Adaptable Platforms

### **Execution Time**





# Example: Smart / Safe City







# Low-Latency, Highest Throughput Inference

30K images / sec on GoogLeNet

AMD EPYC CPU

Alveo U250 Accelerators Performance Scaling Up to 8 Accelerators





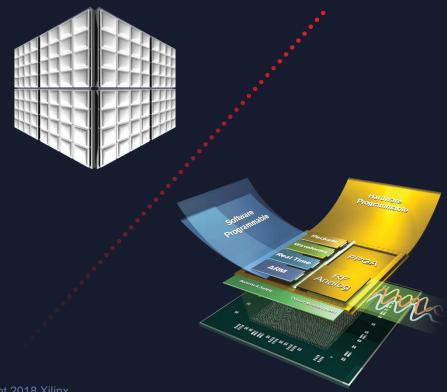
\*batch=1, INT8





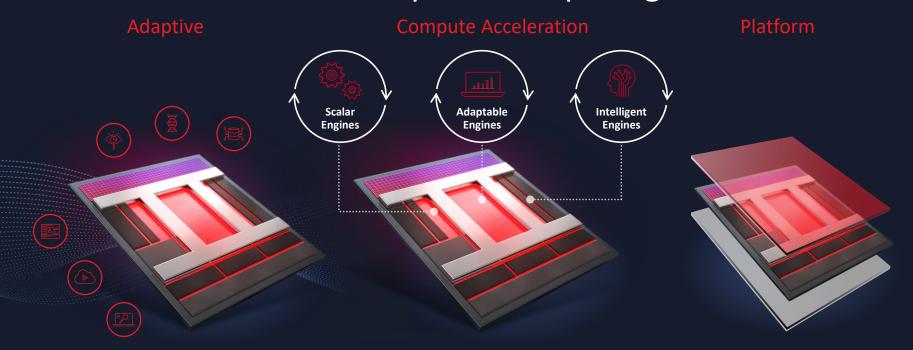
# RFSoC Leading the Way to 5G

- 16nm Zynq MPSoC Architecture with Integrated 4GSPS ADC, 6 GSPS DAC, and Programmable SDFEC
- Enables mMIMO with Dynamic Beamforming
- 7nm Versal with 10x Performance Increase for AI Beamforming and Cell Site Management





# ACAP – The Next Big Advance in Adaptive Computing







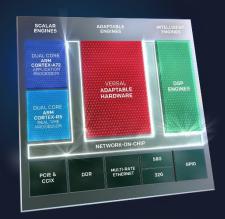




### Versal Prime Series

### **Broad Applicability Across Multiple Markets**

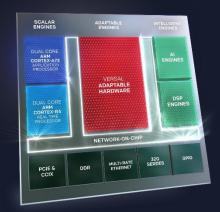
- Mid-Range Series in the Versal Portfolio
- Optimized for Connectivity
- For In-Line Acceleration and Diverse Workloads



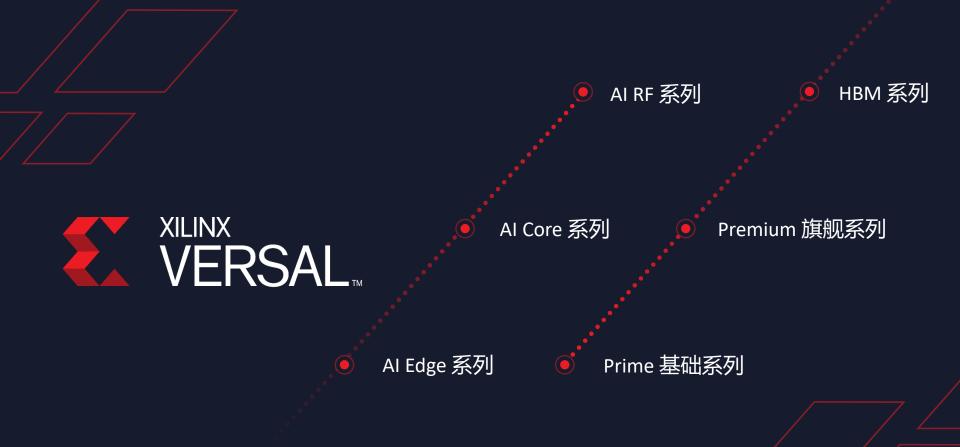
### Versal AI Core Series

### Breakthrough Al Inference Throughput

- Portfolio's Highest Throughput for Low-Latency Inference
- Optimized for Cloud, Networking, and Autonomous Applications
- For Highest Dynamic Range of AI and Workload Acceleration









## Platform for Any Developer

**User Application** C, C++, Python Adaptable for **Any Application** Application-Specific Frameworks Machine Learning | Video | Genomics | Search | Financial Modeling | Database New Unified Software Development Environment Software Programmable Xilinx & Ecosystem **Custom HW** C, Xilinx Libraries OS & Embedded Run-Time **HW Libraries** Scalar Engines Adaptable Engines **Intelligent Engines** Heterogeneous Platform **VERSAL** 



## Low-Latency CNN Inference Performance





# Accelerating AI and Innovation End-to-End in China

Xilinx @ China International Import Expo (CIIE)

Xilinx Developer Forum (XDF)



CIIE Shanghai, Nov. 5-10, 2018

XDF Beijing,Oct.16,2018



Building the Adaptable, Intelligent World

